

Original Paper

Predicting Therapeutic Response to Unfractionated Heparin Therapy: Machine Learning Approach

Ahmad Abdel-Hafez^{1,2}, PhD; Ian A Scott^{3,4*}, MBBS, MHA, MEd; Nazanin Falconer^{5,6*}, PhD; Stephen Canaris^{1*}, BSc; Oscar Bonilla^{1*}, BSc; Sven Marxen^{7*}, BPharm (Hons), Grad Dip Clin Pharm; Aaron Van Garderen^{1,7*}, BPharm (Hons), GCclinPharm, GCProfComp; Michael Barras^{5,8*}, BPharm, Grad Dip Clin Pharm, PhD

¹Clinical Informatics, Metro South Health, Queensland Health, Brisbane, Australia

²School of Public Health & Social Work, Queensland University of Technology, Brisbane, Australia

³Department of Internal Medicine and Clinical Epidemiology, Princess Alexandra Hospital, Brisbane, Australia

⁴Greater Brisbane School of Clinical Medicine, University of Queensland, Brisbane, Australia

⁵Department of Pharmacy, Princess Alexandra Hospital, Brisbane, Australia

⁶Centre for Health Services Research, Faculty of Medicine, The University of Queensland, Brisbane, Australia

⁷Pharmacy Service, Logan and Beaudesert Hospitals, Logan, Australia

⁸School of Pharmacy, University of Queensland, Brisbane, Australia

*these authors contributed equally

Corresponding Author:

Ahmad Abdel-Hafez, PhD

Clinical Informatics, Metro South Health, Queensland Health

199 Ipswich rd

Brisbane, QLD 4102

Australia

Phone: 61 431059039

Email: ahmad.abdel-hafez@health.qld.gov.au

Abstract

Background: Unfractionated heparin (UFH) is an anticoagulant drug that is considered a high-risk medication because an excessive dose can cause bleeding, whereas an insufficient dose can lead to a recurrent embolic event. Therapeutic response to the initiation of intravenous UFH is monitored using activated partial thromboplastin time (aPTT) as a measure of blood clotting time. Clinicians iteratively adjust the dose of UFH toward a target, indication-defined therapeutic aPTT range using nomograms, but this process can be imprecise and can take ≥ 36 hours to achieve the target range. Thus, a more efficient approach is required.

Objective: In this study, we aimed to develop and validate a machine learning (ML) algorithm to predict aPTT within 12 hours after a specified bolus and maintenance dose of UFH.

Methods: This was a retrospective cohort study of 3019 patient episodes of care from January 2017 to August 2020 using data collected from electronic health records of 5 hospitals in Queensland, Australia. Data from 4 hospitals were used to build and test ensemble models using cross-validation, whereas data from the fifth hospital were used for external validation. We built 2 ML models: a regression model to predict the aPTT value after a UFH bolus dose and a multiclass model to predict the aPTT, classified as subtherapeutic (aPTT <70 seconds), therapeutic (aPTT 70-100 seconds), or supratherapeutic (aPTT >100 seconds). Modeling was performed using Driverless AI (H2O), an automated ML tool, and 17 different experiments were iteratively conducted to optimize model accuracy.

Results: In predicting aPTT, the best performing model was an ensemble with 4x LightGBM models with a root mean square error of 31.35 (SD 1.37). In predicting the aPTT class using a repurposed data set, the best performing ensemble model achieved an accuracy of 0.599 (SD 0.0289) and an area under the receiver operating characteristic curve of 0.735. External validation yielded similar results: root mean square error of 30.52 (SD 1.29) for the aPTT prediction model, and accuracy of 0.568 (SD 0.0315) and area under the receiver operating characteristic curve of 0.724 for the aPTT multiclassification model.

Conclusions: To the best of our knowledge, this is the first ML model applied to intravenous UFH dosing that has been developed and externally validated in a multisite adult general medical and surgical inpatient setting. We present the processes of data collection, preparation, and feature engineering for replication.

KEYWORDS

heparin; activated partial thromboplastin time; aPTT; predictive modeling; machine learning; personalized medicine

Introduction

Background

Unfractionated heparin (UFH) is a parenteral anticoagulant used for the prevention and treatment of arterial and venous thromboembolic diseases [1,2]. UFH consists of a heterogeneous mixture of polysaccharides with varying molecular lengths and weights; therefore, direct monitoring of serum drug concentrations to guide optimal dosing is not possible [3,4]. Instead, a surrogate of bleeding time, activated partial thromboplastin time (aPTT), is used to monitor the dose-dependent response [5]. The initial bolus and maintenance doses of UFH are estimated by clinicians using weight-based formulas (units of UFH/kg for bolus and units or UFH kg/hour for maintenance), with the aim of achieving a defined therapeutic aPTT range. Future doses are continually adjusted to maintain this therapeutic range (TR) [6-8], which varies depending on the therapeutic indication [1]. An aPTT value below the TR (subtherapeutic) is linked to reduced efficacy (high probability of recurrence or progression of thromboembolic events), whereas values above the TR (supratherapeutic) are linked to the risk of bleeding [9,10]. For patients with life-threatening thromboembolic events, clinicians aim to rapidly achieve a therapeutic aPTT and maintain a TR for the duration of UFH therapy. In the hospital setting, UFH therapy commences with a bolus (loading) dose followed by a maintenance infusion, and an aPTT is quantified within 12 hours [1,6]. This result provides guidance for further dosing, and clinicians often rely on dosing nomograms ([Multimedia Appendix 1](#)).

UFH is an extremely complex and difficult drug to accurately dose. The UFH molecules are distributed freely throughout the body; bind to many physiological sites including clotting factors, endothelial cells, and macrophages [4]; and are eliminated from the body via several physiological pathways. This creates marked variation in its pharmacokinetics and dose response between patients, such that there is no standardized *one-dose-fits-all* strategy [8-11]. Despite the use of nomograms to optimize dosing, it is difficult to achieve and maintain a TR that places patients at risk. For example, excessive dosing may result in up to 5.5% of patients having a bleeding event [12]. Studies evaluating metrics of safety and effectiveness, such as time to TR, time within TR, and percentage of patients within TR, have demonstrated an inability to predict optimal dosing with confidence [13-15]. The time to TR after initiation of UFH can be as long as 60 hours in some studies, and a recent local study of 200 patients showed a median time to TR of 36 hours [16]. In another study, only 29% of the patients had 2 consecutive therapeutic aPTTs [15] throughout the duration of treatment. Even in large clinical trials with strict patient monitoring, the percentage of patients attaining aPTT in TR within 48 hours is less than 50% [17-19]. Clearly, many factors influence bodily responses to UFH, which are independent of

body weight and are not accounted for in current dosing strategies [7].

Related Work

Machine learning (ML) is a subset of artificial intelligence that identifies patterns in large data sets and encodes them into models to predict new data [20,21]. ML has great potential for providing decision support tools in modern health care [20,22-24], which are developed using large volumes of digitized patient data contained within electronic health records (EHRs) [25-27]. To achieve optimal dosing of UFH, ML methods can potentially be used to develop models that make accurate predictions for the target aPTT in response to UFH dosing. However, there have been few studies to date on how to use ML to optimize UFH dosing [28]. A recent systematic review [28] identified 8 studies using ML for UFH. Out of these, 4 studies predicted aPTT [29-32]; 1 study [33] reported out-of-TR surrogates for aPTT, including bleeding and clotting events; and the remaining 3 studies [34-36] evaluated UFH dosing in hemodialysis patients [28]. To date, 5 studies [29,30,32,33,36] have been conducted in the intensive care units (ICUs) of hospitals in the United States using retrospective data and 3 studies in the dialysis setting [34-36].

A variety of modeling approaches were reported. Four studies reported supervised learning methods including random forests, adaptive boosting, extreme gradient boosting, and neural networks [30,32,34,36]. One study used an unsupervised approach to train the model, which was then fine-tuned using a supervised approach [34]. Three studies also used regression analysis [29,30,34], 2 studies used a reinforcement learning approach to develop their models [33,36], and 1 study [31] compared neural networks with nonlinear mixed-effects modeling methods. Studies have reported a wide range of performance metrics including accuracy, precision, recall, area under the receiver operating characteristic curve (AUC), F_1 -score (a combination of precision and recall), and coincidence rates. The study by Su et al [32] reported the best model accuracy at 88%. Ghassemi et al [29,30] reported 2 studies on modeling for the prediction of UFH dosing, with 1 study reporting a model developed using multinomial regression to predict subtherapeutic and supratherapeutic aPTT, which had superior performance to ML methods. Their later work explored 4 modeling methods, including reinforcement learning and neural networks, with modest accuracies ranging from 0.56 to 0.6. Overall, the multinomial regression model outperformed the ML methods and was a more appealing model because of its clinical interpretability, which is important in the context of implementation and stakeholder engagement. All studies provided limited reporting of reproducibility, and, except for one study by Smith et al [31], none were validated in a new cohort. Most recently, Li et al [37] reported the development and validation of a multiclass aPTT model and subsequent dose prediction application for use in the ICU setting using a shallow neural network approach. The top 5 features for both data sets

included a patient's baseline aPTT, patient weight, total UFH administered, serum creatinine, and age. The model reported performance metrics similar to those of prior studies, with an F_1 -score of 0.887 to 0.925. As with prior studies, the study population was limited to ICU patients and may not be generalizable to other clinical settings. It also does not provide guidance on the exact dose changes that clinicians desire at the point of care.

As evidenced by a recent systematic review [28], no identified models had their impact evaluated within routine clinical practice, and research remains limited and of variable quality. In this body of work, we aimed to develop a model that could be used in hospital general medical and surgical wards, which overcomes the limitations of previous studies with regard to methods, reporting, and external validation.

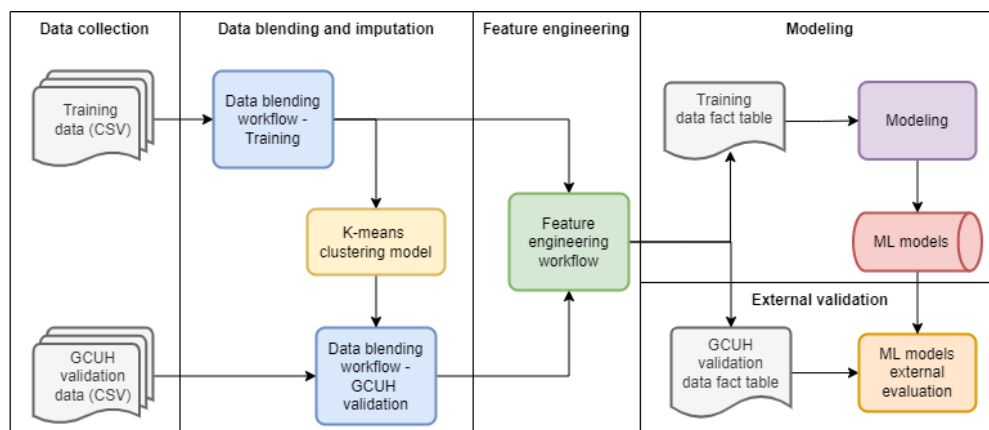
Methods

Data Flow

Figure 1 depicts the data flow and architecture of the project, which is divided into 5 phases. Phase 1 outlines the data

collection in which data files were extracted from EHRs from 5 hospitals, of which 4 were used in model development and the fifth was retained for external validation. In phase 2, the 2 data sets underwent the same phase 2 transformations and mapping process (data blending and imputation), except that the clusters built using the training data set were used unchanged in the validation data set to prevent outcome leakage, where parts of the training data used to create a model were not available at the time of prediction. After data blending, we conducted feature engineering in phase 3, which was again applied to both data sets using the same process. The outputs from the feature-engineering workflow were 2 fact tables: the engineered training data, which were input into the H2O Driverless AI tool to build the ML models (phase 4), and the validation data used to validate the model predictions (phase 5). The same pipeline structure, as shown in Figure 1, was applied in developing and validating a regression model to predict the exact aPTT value and a multiclass model to predict the aPTT class (subtherapeutic, therapeutic, and supratherapeutic), with minor differences in data blending to prepare the outcome columns.

Figure 1. Experiment setup including training and validation processes. GCUH: Gold Coast University Hospital; ML: machine learning; CSV: comma separated values.



Ethics Approval

This research work was granted a low-risk research protocol approval and a waiver of consent by the Metro South Health (MSH) Human Research Ethics Committee for ethical and scientific review (reference number LNR/2019/QMS/54581). We confirm that the work completed in this project is consistent with ethics approval of the acquired research.

Data Collection

EHR data were collected retrospectively for patients admitted between 2017 and 2020 on consecutive admissions to 5 digital hospitals (one health district) in Queensland, Australia, in which UFH was administered for therapeutic purposes. Model development and external validation were undertaken within the Clinical Informatics Division of MSH. We collected data on UFH that were prescribed using a power plan, which is an EHR decision support tool for specific clinical scenarios that facilitates timely ordering of laboratory tests, medication prescribing, and interdisciplinary communication. Four adult-specific power plans that MSH clinicians use, which were

used to identify patients eligible for study inclusion, were acute coronary syndrome, deep vein thrombosis or pulmonary embolism, bridging therapy for oral anticoagulants (warfarin replacement), and low-target-range aPTT anticoagulation for neurosurgical patients. This initial patient cohort was then filtered based on the selection criteria defined by the clinician authors:

- Inclusion criteria: adult patients administered a UFH bolus dose and a maintenance infusion for more than 48 hours, had a documented power plan, and had an aPTT result recorded within 12 hours of the UFH bolus dose.
- Exclusion criteria: ICU patients as ICUs use an ICU-specific EHR that is not linked or integrated into the general EHR system in MSH.

A total of 2783 hospital admissions were identified, involving 2470 patients at the 4 hospitals in MSH whose data were used for model development and 236 hospital admissions involving 221 patients at the hospital where data were used for external validation.

Next, we determined the data tables to be collected from the EHRs and generated an initial list of features. Using previous studies in the literature and the content expertise of collaborating clinicians, we identified 15 data tables that were intentionally inclusive at this stage, while recognizing that some would be removed later if found to be irrelevant or if the data were incomplete. [Multimedia Appendix 2](#) lists the tables showing the number of features extracted from each table before and after the feature-engineering phase.

Data Blending and Imputation

Using the identifier codes of enrolled patients, an aPTT fact table was built by blending UFH bolus dose administration with subsequent aPTT assay results. The rules for inclusion were defined in collaboration between data scientists and clinician researchers to be consistent with the existing literature and to eliminate data noise and ensure data consistency. The following rules were applied:

- A UFH bolus dose was included if it was a de novo (first) dose or was administered after at least 6 hours following prior UFH therapy cessation (equal to approximately 5 UFH elimination half-lives to ensure that no drug remained) [38].
- The aPTT results recorded for the first time after 12 hours of the UFH bolus dose were considered invalid.
- UFH maintenance infusions (maintenance dose) were considered invalid if they were not administered or intravenous infusions were completed, stopped, or paused for more than 1 hour before aPTT testing [39].

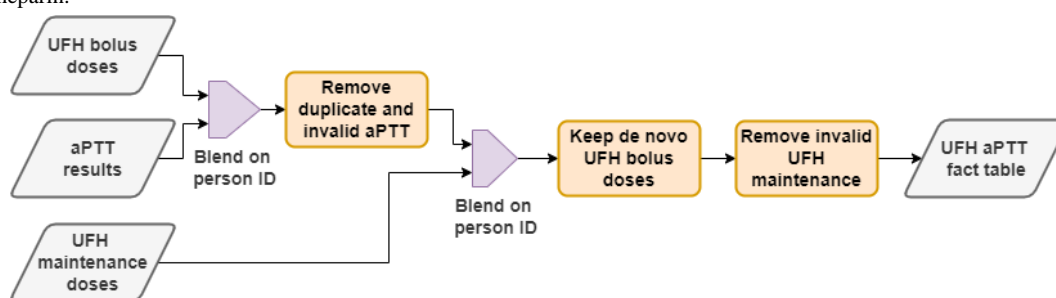
The generation of the aPTT fact table is illustrated in [Figure 2](#). The data blending process was completed in 6 steps, used patient identifiers, and recorded time stamps to connect and filter the data records. The blending process was performed to satisfy the inclusion or exclusion rules previously defined, resulting in a

data set of 2158 records for the model development data set and 236 records for the external validation data set.

During the blending of UFH and aPTT data, several features were identified based on clinician input as listed in [Textbox 1](#). For the other tables ([Figure S1 in Multimedia Appendix 2](#)), we excluded all records documented after the time the target aPTT had been performed, as derived from the fact table. Looking at the counts, we excluded 3 tables as they had an insufficient number of examples to incorporate into the model, with each having less than (278/2783, 9.98%) of the total records in the fact table. During the blending process, we first identified the columns of interest in each of the remaining tables. For some features, such as age and sex, data were added to the aPTT fact table with minor or no processing. Other features, to be useful, required data to be aggregated, grouped, or converted in some way. For example, a less granular mapping was applied to 166 distinct order catalogs of medications to categorize them into medication classes. However, during modeling, only the antimicrobial group was used because of uneven distributions across the data set for the other groups. A complete feature list with details of the applied processing is included in [Table S2 in Multimedia Appendix 2](#).

Dealing with missing data was the next step after blending all the identified tables into a single fact table ([Table 1](#); [Figure S2 in Multimedia Appendix 2](#) provides more detail on all features and definitions). In general, we used clinician expertise to decide on the imputation methods for achieving the most accurate representation of missing values. Imputing the missing baseline aPTT ([Table 1](#)) assumed a normal physiological aPTT value of 30 seconds on the basis of the literature [40] and the median result derived from our training data set. Missing values of patients height and weight were imputed to the mean value of the cohort after grouping by age (bin interval of 10 years), sex, and marital status.

Figure 2. Unfractionated heparin and activated partial thromboplastin time tables blending. aPTT: activated partial thromboplastin time; UFH: unfractionated heparin.



Textbox 1. Features identified during blending of activated partial thromboplastin time (aPTT) and unfractionated heparin (UFH) administration data.

Feature and description
• Baseline aPTT: aPTT result preceding the current (target) aPTT
• Baseline aPTT minutes: time (in) between the baseline aPTT and target aPTT
• UFH bolus minutes: time (in) between the bolus dose and target aPTT
• UFH maintenance minutes: time (in minutes) between the maintenance start and target aPTT

Table 1. Missing data handling.

Features	Imputation
Baseline aPTT ^a and baseline aPTT minutes	Missing values and values completed more than 24 hours before UFH ^b bolus administration were imputed to 30 seconds, whereas baseline aPTT minutes are imputed to 1440 (24 hours) minutes for those records.
Weight and height	Encounters with no measurements were imputed to the averages for their age, marital status, and sex (as recorded in the patient electronic health record).
Vital Signs features	If the results are missing or occurred more than 12 hours before the target aPTT, they are imputed using centroid values of k-means clustering with k=10.
Pathology results	If the results are missing or occurred more than 1 week before the target aPTT, they are imputed using k-means clustering using centroid value with k=10.
Waterflow score and ADLs ^c	Imputed to 0 where missing or older than 1 week before target aPTT.

^aaPPT: activated partial thromboplastin time.

^bUFH: unfractionated heparin.

^cADL: activity of daily living.

Feature Engineering and Data Transformation

In this phase, the blended and imputed aPTT fact table was used initially to conduct univariate analysis and data visualization, which aimed to inform decisions about building new features and transforming data. However, this process was not separate from data modeling; rather, it was an iterative process where ML models were built on initial features that changed and evolved, thus serving as new feature inputs to the next cycle of modeling. [Multimedia Appendix 3](#) provides details and visualizations of Pearson correlations between features and outcomes in our aPTT fact table.

[Table 2](#) summarizes the demographic data and important features that are most relevant to the blended (training) data set. The definitions of diagnoses were based on the International Classification of Diseases (ICD)-10 codes; however, we were only able to include categories with large frequencies; that is, ACS and VTE. Other diagnoses were grouped as other. All patients' recorded codes during their admission were used in the grouping process.

The reported aPTT result showed a distribution heavily skewed to the right ([Figure 3](#)) and contained outliers, which negatively impacted the performance of a regression model. Although several statistical methods, such as quadratic mean learning [41], can be used to correct this, we chose, on the basis of clinician expertise, to reduce the negative impact of skewness by introducing a floor and ceiling value to target aPTT of 30 and 150. Values less than 30 seconds reflect normal physiological values. The impact of using floor and ceiling values is visualized in residual graphs in [Figure S1](#) in [Multimedia Appendix 3](#), and more feature analysis information is presented in [Figures S3-S5](#) in [Multimedia Appendix 2](#).

Four calculated features were introduced. The first one was the UFH maintenance dose where, unlike the single bolus dose administration, the cumulative maintenance dose was derived based on the total units in the syringe, the infusion period, and the total infusion time before the target aPTT test was performed, excluding any stoppage periods of the infusion (calculation described as follows):

$$UFH_{Maintenance} = (UFH \text{ syringe size} / \text{Total infusion period}) \times (\text{infusion time} - \text{infusion stop})$$

The standard amount of UFH contained in a syringe was 25,000 units (50 mL syringe, 500 IU/mL), and this, together with the total time for the syringe to be emptied with no interruption, indicated the infusion rate as the number of UFH units infused per minute. The second part of the equation (infusion time—infusion stop) aimed to calculate the exact infusion period in minutes. The 3 other calculated features were body size, UFH bolus time, and UFH bolus time; body size is calculated using the following equations:

$$\text{Size} = \text{Weight} / \text{Height}$$

$$\text{UFH Bolus Time} = \text{UFH Bolus Dose} / \text{Time to aPTT}$$

$$\text{UFH Bolus Time and Size} = \text{Size} \times \text{UFH Bolus Time}$$

Finally, we added a cosine cyclical transformation of aPTT time to build 3 features representing the aPTT day of the week, hour of the day, and month of the year. At the end of this phase, we obtained 93 features in the aPTT fact table. Depending on the data distributions, continuous variables were scaled using the Yeo-Johnson transformation [42] from the SciPy Python library [40] or a min-max transformation. Details about the transformation method applied for every feature are provided in [Figure S2](#) in [Multimedia Appendix 2](#), and equation details are available in [Multimedia Appendix 4](#).

Table 2. Baseline characteristics of training data set (all records, N=2783).

Feature	Values
Sex, n (%)	
Male	1898 (68.2)
Female	885 (31.8)
Diagnosis, n (%)	
ACS ^a	818 (29.4)
VTE ^b	540 (19.4)
Age (years), mean (SD)	65.8 (14.6)
Weight (kg), mean (SD)	87.8 (26.7)
Baseline aPTT ^c (seconds), mean (SD)	36 (11.1)
UFH ^d bolus dose (units), mean (SD)	4713 (1467)
UFH maintenance (units), mean (SD)	6767 (4993)
Time between UFH bolus and aPTT (minutes), mean (SD)	364.1 (149)

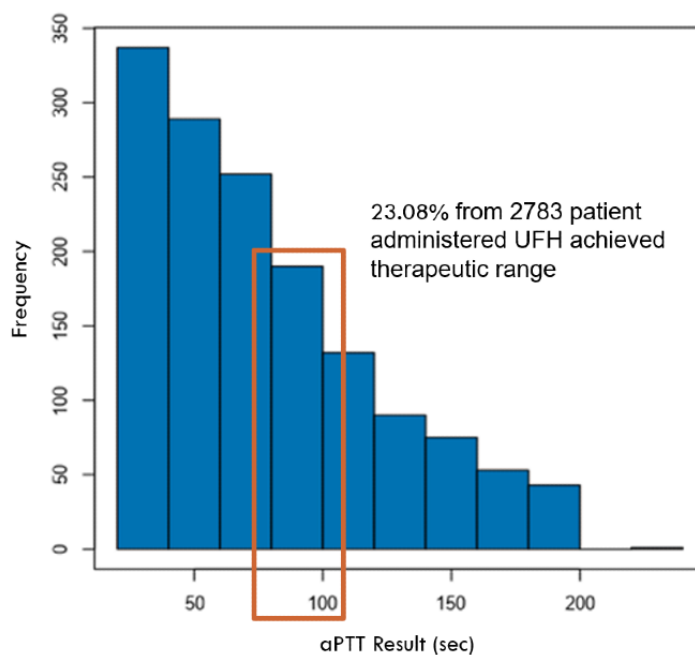
^aACS: acute coronary syndrome.

^bVTE: venous thromboembolism.

^caPTT: activated partial thromboplastin time.

^dUFH: unfractionated heparin.

Figure 3. Frequencies of target activated partial thromboplastin time results with bin size=20. aPTT: activated partial thromboplastin time.



Modeling

Outcomes and Setup

In this phase, 2 models were developed: a regression model for predicting the target aPTT result and a multiclassification model for predicting the aPTT class as subtherapeutic, therapeutic, or supratherapeutic. To identify the optimal model in each case, several models were iteratively tested, with each iteration evaluated using 3-fold cross-validation involving 67 by 33 data splits, where all cases could be used for both model training and internal validation. Cross-validation was repeated 3 times

to ensure that the validation metrics were robust, as the training data sets were relatively small. The predictive metrics of all iterations were averaged to obtain the overall results for the model.

The modeling process was completed using the H2O Driverless AI tool, which is an auto-ML tool that takes tabular data as input and builds supervised models automatically using the available open-source ML libraries in Python and R. It also automates model validation, tuning, and selection to achieve an accuracy level equivalent to that of the manually built models. The tool also performs an iterative feature evolution process during

modeling to discover new features. Supervised ML models supported by H2O Driverless AI include XGBoost [43], LightGBM [44], generalized linear models [45], TensorFlow [42], RuleFit [46], and FTRL [47] (followed by the regularized leader) [48]. The tool will generate and test large numbers of models by using different open-source algorithms, undertake hyperparameter tuning, try different feature subsets, and combine models using different methods. After generating hundreds of different models and combinations (ensembles), the tool recommended the most accurate model built for deployment.

Ensemble Regression Model

We built an ensemble regression model to predict aPTT values within 12 hours of a UFH bolus dose. The optimized performance metric was the root mean square error (RMSE). Other reported metrics include the mean absolute error and coefficient of determination (R^2).

During this experiment, 1126 alternative models were trained, including constant predictions, the LightGBM [44] and XGBoost [43] algorithms, and ensemble models. After the feature evolution process, the 93 original features were converted into 188 features, with the contributing features on every model automatically selected by the H2O Driverless AI tool during the training process. We built several baseline regression models, against which the performance of the H2O Driverless AI ensemble model was compared. For all these baseline models, we used the same set of features used to build the ensemble model, except for those evolved during the modeling process using the H2O Driverless AI tool. The first baseline model was built using the tool but as a single model rather than an ensemble. The best model returned by the tool is the XGBoost model. The other baseline models were developed using the Python scikit-learn library, where we tested 3 different linear regression models: Ridge [49], Lasso [50], and ElasticNet [51].

Ensemble Multiclassification Model

We built a multiclassification model using the same training data set to predict the target aPTT class, where aPTT < 70 seconds was considered subtherapeutic, aPTT between 70 and 100 seconds as therapeutic, and aPTT > 100 seconds as suprathematic. In this modeling process, we optimized the accuracy and reported several other metrics relevant to multiclassification, including macroprecision, macrorecall, macro- F_1 -score, and macroaverage one class vs rest classes AUCs [52].

In total, 457 different models were trained and tested using the H2O Driverless AI tool. Similar to the regression models, the tool tested constant predictions, the LightGBM [44] and

XGBoost [43] algorithms, and ensemble models. The evolved and original features used in the modeling process totaled 2196 features. The tool ranked the models based on their performance and recommended the best performing model producing the best accuracy. The other baseline models were developed using the Python scikit-learn library: logistic regression, logistic regression with recursive feature elimination, support vector machine [53] using a linear support vector classifier [54], and support vector machine using polynomial support vector classifier.

External Validation

External validation was performed using data obtained from patient records at a fifth hospital (Gold Coast University Hospital), which had an exact schema and table structure as the training set. The final data set comprised 236 records, after applying the same inclusion or exclusion criteria as the training data set. We used the method proposed by Archer et al [55] to calculate the sample size sufficient to validate the proposed model and achieve a preselected target for the CI of R^2 . The equations and calculation details are provided in [Multimedia Appendix 4](#). The equation generated a minimum sample size of at least 235 participants, which was achieved.

Results

Ensemble Regression Model

The best performing model was an ensemble of 4x LightGBM models that were linearly blended. LightGBM is a gradient-boosting framework that uses tree-based learning algorithms [44]. The model relied on 134 features, some of which are well-established as influencing responses to UFH, such as weight and baseline aPTT, with others first identified in this experiment, of which the most important was the time between when the bolus dose was administered and when the aPTT was measured (Table 3). The bolus dose time, baseline aPTT, age, and bolus dose were also relatively important. Weight, size (weight divided by height), and hematological and biochemical parameters, including serum creatinine level, as a measure of renal function, were also among the top 20 features. Table 4 shows the description and selected list of the LightGBM hyperparameters.

The performance metrics of our ensemble and all the baseline models are listed in Table 5. The H2O Driverless AI ensemble model had best performance with RMSE 31.35 (SD 1.37), residual charts provided in Figure S1 in [Multimedia Appendix 3](#). In addition, this baseline model outperformed all other Python-based linear regression models because the tool tested several algorithms, as previously mentioned, and evolved additional features during the modeling process.

Table 3. Top 10 most important features with relative importance scores.

Feature	Relative importance
Minutes between UFH ^a bolus and aPTT ^b	1
UFH bolus time	0.58
Baseline aPTT	0.41
Age	0.37
UFH bolus dose	0.3
Height	0.25
UFH maintenance	0.23
UFH bolus size calculated	0.22
Weight	0.21
Size calculated	0.195

^aUFH: unfractionated heparin.

^baPTT: activated partial thromboplastin time.

Table 4. Descriptions of contributing models in the final regression ensemble model.

ID	Model type	Model weight	Fitted features	Feature fraction	Max leaves	Learning rate	Max bins	Lambda L1	Lambda L2
0	LightGBM	0.3333	65	0.6	16	0.01	128	0	0.5
1	LightGBM	0.1042	69	0.6	16	0.01	128	0	0.5
2	LightGBM	0.1667	95	0.8	64	0.01	256	0	10
3	LightGBM	0.3958	134	0.4	16	0.01	64	0	2

Table 5. Performance of regression models for predicting activated partial thromboplastin time results.

Tool	Model	Mean absolute error	Root mean square error	R^{2a}
H2O DAI ^b	Final ensemble model	24.61 ^c	31.35	0.355
H2O DAI	XGBoost	25.51	32.33	0.31
SKlearn ^d	Linear regression	26.89	33.8	0.244
SKlearn	Ridge regression	26.93	33.79	0.244
SKlearn	Lasso regression	26.93	33.68	0.249
SKlearn	Elasticnet regression	27.15	33.72	0.247

^a R^2 coefficient of determination.

^bDAI: Driverless AI.

^cMinimum error rate.

^dSKlearn: a machine learning library in Python.

Ensemble Multiclassification Model

The best performing model was a linear blend ensemble of 4 different models with different weights, 2 XGBoost models and 2 LightGBM models (Table 6).

The performance metrics of the ensemble multiclassification model and baseline models built using the SKlearn library in Python are presented in Table 7, with the ensemble model showing superior performance across all metrics, with an accuracy of 0.599 and macro- F_1 -score of 0.613. The simple logistic regression model in Python was the second-best performer, highlighting the efficiency of using auto-ML tools

for feature engineering, evolution, and model tuning and blending.

Figure 4 shows the confusion matrix for the ensemble model, demonstrating very good accuracy (0.88) for the subtherapeutic class aPTT<70 seconds, intermediate accuracy (0.512 for the supratherapeutic class aPTT>100 seconds, and poor accuracy (0.098) for the therapeutic class aPTT 70 to 100 seconds. This lower accuracy is most likely a result of class imbalance due to the underrepresentation of the therapeutic class in the data set. Nevertheless, predicting patients at risk of recurrent thromboembolic events from underdosing or at risk of bleeding from overdosing is important for clinicians and patients.

For the multiclassification ensemble model, the validation set returned an accuracy of 0.568 (95% CI 0.538-0.598) and an AUC of 0.724 (95% CI 0.714-0.734), which also compares favorably with the corresponding values for the training set cross-validation of 0.599 and 0.735, respectively. In surveying

the confusion matrix (Figure 5), the model demonstrated similar accuracy for each class as the training model: 0.899 for the aPTT class <70 seconds, 0.492 for the aPTT class >100 seconds, and 0.078 for the aPTT class 70 to 100 seconds.

Table 6. Descriptions of contributing models in the final multiclassification ensemble model.

ID	Model type	Model weight	Fitted features	Feature fraction	Max leaves	Learning rate	Max bins	Lambda L1	Lambda L2
0	XGBoost	0.3067	1900	0.2	8	0.01	128	0	0.5
1	XGBoost	0.2	1914	0.5	8	0.01	256	0	5
2	LightGBM	0.4	78	0.6	16	0.01	64	0	0.5
3	LightGBM	0.0933	183	0.8	64	0.01	256	0	0

Table 7. Performance of multiclassification models in predicting activated partial thromboplastin time class.

Tool	Model	Accuracy	Macroprecision	Macrorecall	Macro-F ₁ -score	Macro-AUC ^a
H2O DAI ^b	Final ensemble model	0.599 ^c	0.554	0.686	0.613	0.735
SKlearn	Logistic regression	0.562	0.51	0.56	0.52	0.691
SKlearn	Logistic regression with RFE ^d	0.557	0.49	0.56	0.5	0.687
SKlearn	SVM ^e —linear SVC ^f	0.535	0.51	0.54	0.517	0.679
SKlearn	SVM—polynomial SVC	0.451	0.46	0.45	0.457	0.614

^aAUC: area under the receiver operating characteristic curve.

^bDAI: Driverless AI.

^cBest calculated accuracy.

^dRFE: recursive feature elimination.

^eSVM: support vector machine.

^fSVC: support vector classifier.

Figure 4. Multiclassification confusion matrix. aPTT: activated partial thromboplastin time.

		Predicted			
		aPTT<70	100≥aPTT≥70	aPTT>100	Total
Actual	aPTT<70	1216	32	134	1382
	100≥aPTT≥70	381	63	199	643
	aPTT>100	324	46	388	758
	Total	1921	141	721	2783

Figure 5. Multiclassification confusion matrix for external validation. aPTT: activated partial thromboplastin time.

		Predicted			
		aPTT<70	100≥aPTT≥70	aPTT>100	Total
Actual	aPTT<70	98	2	9	109
	100≥aPTT≥70	30	5	29	64
	aPTT>100	31	1	31	63
	Total	159	8	69	236

Discussion

Principal Findings

This study reports the development and external validation of an ML model for the prediction of aPTT following bolus and maintenance dosing with UFH. The ML models were developed using EHR data from 4 Australian hospitals with the best performing approach, producing an ensemble with 4x LightGBM models with an RMSE of 31.35. As a multiclassification task, the ensemble model achieved an accuracy of 0.599 and an AUC of 0.735. External validation in a new patient cohort at a fifth hospital showed similar results, with an RMSE of 30.52 for the prediction model and an accuracy of 0.568 and AUC of 0.724 for the multiclassification model.

The final model relied on 93 features, including body weight, baseline aPTT, and bolus dose, with others novel to this study and contemporary clinical knowledge, such as hematological and biochemical features (Multimedia Appendices 2 and 3). The most important clinically informative novel features were the time between administration of the bolus dose and aPTT, age, and baseline aPTT. Baseline aPTT, maintenance UFH dose, and time between bolus administration and aPTT as a grouped feature, which had the highest relative importance (Table 3). UFH is considered a high-risk drug with a narrow therapeutic window, and therefore requires patient-specific dosing to ensure safety and effectiveness [1,7]. Determining the optimal initial bolus and maintenance dosing for UFH therapy is challenging because of the many unknown physiological variables that may contribute to its anticoagulant response. Initial bolus dosing based on body weight is currently preferred [19]; however, other variables must influence the response [56,57]. Nomograms and regular aPTT monitoring, which guide subsequent dose adjustments, increase the proportion of patients achieving a target therapeutic aPTT range [6,8]. Unfortunately, local data derived from 2783 patient episodes suggest that this target is achieved in as few as 23.08% of the patients administered UFH.

As UFH is difficult to administer, new anticoagulants have been introduced in the health setting. Although these new anticoagulants, such as direct-acting oral anticoagulants and low-molecular-weight UFHs, have similar effectiveness to UFH in thromboembolic disease, UFH retains an extensive role in hospital practice because of its several advantages [1,2]. Current dosing is based on nomograms, drug action can be quickly reversed if required, the response can be monitored using aPTT, and its short half-life ensures that the drug is quickly eliminated if urgent surgery is required, or bleeding occurs. As per our data, UFH is still a commonly used drug that requires better dosing and monitoring to ensure patient safety than what is currently achievable. Using ML to derive a predictive model offers a possible approach to more accurately predict individual responses to an initial bolus dose of UFH. This information will assist clinicians in estimating the optimal bolus dose. Developing, testing, and deploying these models is becoming more feasible with the advent of large, digitized data sets such as EHRs [22,26,58]; systems that have been implemented in most large hospitals in Australia. Our study demonstrates that

many other features exist beyond the traditional weight-based calculations to determine the best UFH bolus dose. This has the potential to improve the safety profile of UFH. EHR data now afford the opportunity to start using model-based dosing strategies and the ability to develop continuous learning ML models in the future [59].

ML is increasingly being used in early phase drug development [22,26,58] and postmarketing dose design, particularly for other high-risk drugs with narrow therapeutic windows, such as warfarin [56,57,60,61], insulin [62], digoxin [63], immunosuppressants, and chemotherapeutics [64]. Using ML models to guide dosing of UFH in acutely ill, unstable medical and surgical patients to minimize thromboembolic events and bleeding events is an important advancement. In developing such models, as shown in our study, a collaborative approach whereby clinicians and data informaticians work in close consultation is essential. Our study used researchers, data engineers, hematologists, pharmacists, and medical practitioners in its design and conduct. This is essential for developing usable artificial intelligence solutions in hospital settings [65].

Comparison With Prior Work

In this study, an ensemble approach with supervised learning was used. Five other studies have reported using supervised learning in developing models to assist with UFH dosing [29,30,32,34,36]. To date, although 3 report accuracy [32,33,37] superior to that of our ensemble approach, these models were restricted to ICU data sets from the United States and China and are, therefore, not generalizable to the general medical and surgical wards of hospitals where UFH is most frequently administered. Furthermore, compared with all existing studies of ML in UFH dosing, ours was the only one, apart from one small external validation in a hemodialysis setting [31,66], to evaluate model performance when applied to new unseen data. External validation is considered an essential step before assessing the efficacy in controlled clinical trials and subsequent implementation in routine practice [65,67].

Future Work

The stage is now set for a feasibility study, the implementation of the model in hospital clinical workflows, and, if successful, further evaluation of clinical utility in a trial comparing current standard practice with model-guided bolus dosing. Implementing the model in routine practice requires an easily accessible decision support platform that can prepopulate most, if not all, the features within the model from the EHR without the need for manual input by clinicians. The model will need to rapidly provide guidance at the exact time of decision-making and will not require end users to undertake extensive training in its use [68,69].

Limitations

Our model was developed and validated using data from EHRs of 5 hospitals and, therefore, should be tested in other health care systems that use EHRs. The modeling approach only applies to adult inpatients admitted to general medical and surgical specialties. ICU patients were excluded from this study. Furthermore, this modeling approach was intended for the prediction of aPTT after a prespecified bolus and maintenance

dose, and as such, further work is required to allow dose calculation and adjustment.

Some features (such as activities of daily living assessments), which were included in the 93 influential features, may not always be available at the time of dosing UFH, and appropriate surrogates should be considered in future iterations of the model. It is also important to consider the level of data standardization within the EHR data sets, which may limit the applicability and usefulness of ML-derived models [70]. For example, differences in how features are measured (eg, the weight and height using different scales), differences in aPTT assays, or different locations of data in EHR may affect model performance and generalizability.

Finally, similar to many other dosing regimens for intravenous drugs, a perfect algorithm for UFH is probably not achievable, as UFH interacts with a myriad of hematological and physiological factors that may affect its anticoagulant effect. Many of these cannot be measured or remain unknown. The goal of our study was to produce and validate a predictive model

for UFH dosing that is significantly more accurate than the current weight-based nomograms that have been in use for many years.

Conclusions

This study reports the development and validation of an auto-ML-built ensemble modeling approach for predicting aPTT results and determining their therapeutic classification within 12 hours of administration of a de novo UFH bolus accompanied by a UFH maintenance infusion. ML models were developed using retrospective data from the EHRs of the 4 hospitals. These models were shown to have a consistent performance when applied to an external data set from a fifth hospital. To our knowledge, this is the first study of ML regression and multiclassification models applied to UFH dosing that has used auto-ML tools in model development and conducted external validation. Future work should include the optimization of model performance and its redesign and incorporation into a dose calculation software tool that can be easily used by clinicians at the point of care.

Acknowledgments

The authors would like to acknowledge the contributions of different parties in this project, which led directly to its success: Princess Alexandra Hospital, Brisbane, Queensland, Australia and the clinical informatics team; Gold Coast University Hospital: Trudy Teasdale (Assistant Director of Pharmacy) and Liz Coombes (Director of Pharmacy); and MIP Inc, Australia: Jason Burke (Artificial Intelligence and Machine Learning Practice Lead) and Ryan Lambert (Chief Analytics Officer).

Funding for this study was provided by a Metro South Health Research Foundation grant (RSS_2019_024).

Data Availability

Access to the data used in this project is restricted to the research team and for the duration of the research project, as outlined in the national statement on ethical conduct in human research 2007, the Australian code for the responsible conduct of research, and Metro South research management policy and procedures. Access to data was granted by Metro South Health Research Governance Committee, Site Specific Assessment reference number SSA/19/QMS/54581.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Dosing and monitoring nomogram used in Queensland, Australia. Retrieved from “State of Queensland (Queensland Health). Heparin intravenous infusion order and administration form for adults. 11th ed. Queensland 2018.”.

[\[PDF File \(Adobe PDF File\), 851 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Data analysis, description, and processing extra tables.

[\[PDF File \(Adobe PDF File\), 279 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Data visualizations, heatmaps, plots, and residual graphs.

[\[PDF File \(Adobe PDF File\), 492 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Metrics equations.

[\[PDF File \(Adobe PDF File\), 157 KB-Multimedia Appendix 4\]](#)

References

1. Kearon C, Akl EA, Ornelas J, Blaivas A, Jimenez D, Bounameaux H, et al. Antithrombotic therapy for VTE disease: CHEST guideline and expert panel report. *Chest* 2016 Feb;149(2):315-352. [doi: [10.1016/j.chest.2015.11.026](https://doi.org/10.1016/j.chest.2015.11.026)] [Medline: [26867832](https://pubmed.ncbi.nlm.nih.gov/26867832/)]
2. Leentjens J, Peters M, Esselink AC, Smulders Y, Kramers C. Initial anticoagulation in patients with pulmonary embolism: thrombolysis, unfractionated heparin, LMWH, fondaparinux, or DOACs? *Br J Clin Pharmacol* 2017 Nov;83(11):2356-2366 [FREE Full text] [doi: [10.1111/bcp.13340](https://doi.org/10.1111/bcp.13340)] [Medline: [28593681](https://pubmed.ncbi.nlm.nih.gov/28593681/)]
3. Kandrotas RJ. Heparin pharmacokinetics and pharmacodynamics. *Clin Pharmacokinet* 1992 May;22(5):359-374. [doi: [10.2165/00003088-199222050-00003](https://doi.org/10.2165/00003088-199222050-00003)] [Medline: [1505142](https://pubmed.ncbi.nlm.nih.gov/1505142/)]
4. Nutescu EA, Burnett A, Fanikos J, Spinler S, Wittkowsky A. Pharmacology of anticoagulants used in the treatment of venous thromboembolism. *J Thromb Thrombolysis* 2016 Jan;41(1):15-31 [FREE Full text] [doi: [10.1007/s11239-015-1314-3](https://doi.org/10.1007/s11239-015-1314-3)] [Medline: [26780737](https://pubmed.ncbi.nlm.nih.gov/26780737/)]
5. Smythe MA, Koerber JM, Westley SJ, Nowak SN, Begle RL, Balasubramaniam M, et al. Use of the activated partial thromboplastin time for heparin monitoring. *Am J Clin Pathol* 2001 Jan;115(1):148-155. [doi: [10.1309/N7RA-D6XN-9EQX-ABJ9](https://doi.org/10.1309/N7RA-D6XN-9EQX-ABJ9)] [Medline: [11190801](https://pubmed.ncbi.nlm.nih.gov/11190801/)]
6. Raschke RA, Reilly BM, Guidry JR, Fontana JR, Srinivas S. The weight-based heparin dosing nomogram compared with a "standard care" nomogram. A randomized controlled trial. *Ann Intern Med* 1993 Nov 01;119(9):874-881. [doi: [10.7326/0003-4819-119-9-199311010-00002](https://doi.org/10.7326/0003-4819-119-9-199311010-00002)] [Medline: [8214998](https://pubmed.ncbi.nlm.nih.gov/8214998/)]
7. Smythe MA, Priziola J, Dobesh PP, Wirth D, Cuker A, Wittkowsky AK. Guidance for the practical management of the heparin anticoagulants in the treatment of venous thromboembolism. *J Thromb Thrombolysis* 2016 Jan;41(1):165-186 [FREE Full text] [doi: [10.1007/s11239-015-1315-2](https://doi.org/10.1007/s11239-015-1315-2)] [Medline: [26780745](https://pubmed.ncbi.nlm.nih.gov/26780745/)]
8. Zimmermann AT, Jeffries WS, McElroy H, Horowitz JD. Utility of a weight-based heparin nomogram for patients with acute coronary syndromes. *Intern Med J* 2003;33(1-2):18-25. [doi: [10.1046/j.1445-5994.2002.00297.x](https://doi.org/10.1046/j.1445-5994.2002.00297.x)] [Medline: [12534874](https://pubmed.ncbi.nlm.nih.gov/12534874/)]
9. Anand SS, Yusuf S, Pogue J, Ginsberg JS, Hirsh J, Organization to Assess Strategies for Ischemic Syndromes Investigators. Relationship of activated partial thromboplastin time to coronary events and bleeding in patients with acute coronary syndromes who receive heparin. *Circulation* 2003 Jun 17;107(23):2884-2888. [doi: [10.1161/01.CIR.0000077530.53367.E9](https://doi.org/10.1161/01.CIR.0000077530.53367.E9)] [Medline: [12796128](https://pubmed.ncbi.nlm.nih.gov/12796128/)]
10. Hull RD, Raskob GE, Brant RF, Pineo GF, Valentine KA. Relation between the time to achieve the lower limit of the APTT therapeutic range and recurrent venous thromboembolism during heparin treatment for deep vein thrombosis. *Arch Intern Med* 1997;157(22):2562-2568. [Medline: [9531224](https://pubmed.ncbi.nlm.nih.gov/9531224/)]
11. Raschke RA, Gollihare B, Peirce JC. The effectiveness of implementing the weight-based heparin nomogram as a practice guideline. *Arch Intern Med* 1996;156(15):1645-1649. [Medline: [8694662](https://pubmed.ncbi.nlm.nih.gov/8694662/)]
12. Juergens CP, Semsarian C, Keech AC, Beller EM, Harris PJ. Hemorrhagic complications of intravenous heparin use. *Am J Cardiol* 1997 Jul 15;80(2):150-154. [doi: [10.1016/s0002-9149\(97\)00309-3](https://doi.org/10.1016/s0002-9149(97)00309-3)] [Medline: [9230150](https://pubmed.ncbi.nlm.nih.gov/9230150/)]
13. Raschke R, Hirsh J, Guidry JR. Suboptimal monitoring and dosing of unfractionated heparin in comparative studies with low-molecular-weight heparin. *Ann Intern Med* 2003 May 06;138(9):720-723. [doi: [10.7326/0003-4819-138-9-200305060-00008](https://doi.org/10.7326/0003-4819-138-9-200305060-00008)] [Medline: [12729426](https://pubmed.ncbi.nlm.nih.gov/12729426/)]
14. Ting C, Sylvester KW, Schurr JW. Time in the therapeutic range for assessing anticoagulation quality in patients receiving continuous unfractionated heparin. *Clin Appl Thromb Hemost* 2018 Dec;24(9_suppl):178S-181S [FREE Full text] [doi: [10.1177/1076029618798944](https://doi.org/10.1177/1076029618798944)] [Medline: [30213200](https://pubmed.ncbi.nlm.nih.gov/30213200/)]
15. Hylek EM, Regan S, Henault LE, Gardner M, Chan AT, Singer DE, et al. Challenges to the effective use of unfractionated heparin in the hospitalized management of acute thrombosis. *Arch Intern Med* 2003 Mar 10;163(5):621-627. [doi: [10.1001/archinte.163.5.621](https://doi.org/10.1001/archinte.163.5.621)] [Medline: [12622610](https://pubmed.ncbi.nlm.nih.gov/12622610/)]
16. George C, Barras M, Coombes J, Winkel K. Unfractionated heparin dosing in obese patients. *Int J Clin Pharm* 2020 Apr;42(2):462-473. [doi: [10.1007/s11096-020-01004-5](https://doi.org/10.1007/s11096-020-01004-5)] [Medline: [32140914](https://pubmed.ncbi.nlm.nih.gov/32140914/)]
17. Antman EM, TIMI 11B. Enoxaparin versus unfractionated heparin for unstable angina or non-Q-wave myocardial infarction: a double-blind, placebo-controlled, parallel-group, multicenter trial. Rationale, study design, and methods. *Am Heart J* 1998 Jun;135(6):S353-S360. [doi: [10.1016/s0002-8703\(98\)70265-0](https://doi.org/10.1016/s0002-8703(98)70265-0)] [Medline: [9628449](https://pubmed.ncbi.nlm.nih.gov/9628449/)]
18. Cohen M, Demers C, Gurfinkel EP, Turpie AG, Fromell GJ, Goodman S, et al. A comparison of low-molecular-weight heparin with unfractionated heparin for unstable coronary artery disease. Efficacy and safety of subcutaneous enoxaparin in non-q-wave coronary events study group. *N Engl J Med* 1997 Aug 14;337(7):447-452. [doi: [10.1056/NEJM199708143370702](https://doi.org/10.1056/NEJM199708143370702)] [Medline: [9250846](https://pubmed.ncbi.nlm.nih.gov/9250846/)]
19. Cuker A. Unfractionated heparin for the treatment of venous thromboembolism: best practices and areas of uncertainty. *Semin Thromb Hemost* 2012 Sep;38(6):593-599. [doi: [10.1055/s-0032-1319770](https://doi.org/10.1055/s-0032-1319770)] [Medline: [22718256](https://pubmed.ncbi.nlm.nih.gov/22718256/)]
20. Deo RC. Machine learning in medicine. *Circulation* 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.115.001593](https://doi.org/10.1161/CIRCULATIONAHA.115.001593)] [Medline: [26572668](https://pubmed.ncbi.nlm.nih.gov/26572668/)]
21. Scott IA. Demystifying machine learning: a primer for physicians. *Intern Med J* 2021 Sep 19;51(9):1388-1400. [doi: [10.1111/imj.15200](https://doi.org/10.1111/imj.15200)] [Medline: [33462882](https://pubmed.ncbi.nlm.nih.gov/33462882/)]

22. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019 Apr 04;380(14):1347-1358. [doi: [10.1056/NEJMra1814259](https://doi.org/10.1056/NEJMra1814259)] [Medline: [30943338](https://pubmed.ncbi.nlm.nih.gov/30943338/)]
23. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA* 2018 Apr 03;319(13):1317-1318. [doi: [10.1001/jama.2017.18391](https://doi.org/10.1001/jama.2017.18391)] [Medline: [29532063](https://pubmed.ncbi.nlm.nih.gov/29532063/)]
24. Scott I, Cook D, Coiera E. Evidence-based medicine and machine learning: a partnership with a common purpose. *BMJ Evid Based Med* 2021 Dec;26(6):290-294. [doi: [10.1136/bmjebm-2020-111379](https://doi.org/10.1136/bmjebm-2020-111379)] [Medline: [32816901](https://pubmed.ncbi.nlm.nih.gov/32816901/)]
25. Luz CF, Vollmer M, Decruyenaere J, Nijsten MW, Glasner C, Sinha B. Machine learning in infection management using routine electronic health records: tools, techniques, and reporting of future technologies. *Clin Microbiol Infect* 2020 Oct;26(10):1291-1299 [FREE Full text] [doi: [10.1016/j.cmi.2020.02.003](https://doi.org/10.1016/j.cmi.2020.02.003)] [Medline: [32061798](https://pubmed.ncbi.nlm.nih.gov/32061798/)]
26. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. *N Engl J Med* 2016 Sep 29;375(13):1216-1219 [FREE Full text] [doi: [10.1056/NEJMp1606181](https://doi.org/10.1056/NEJMp1606181)] [Medline: [27682033](https://pubmed.ncbi.nlm.nih.gov/27682033/)]
27. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc* 2017 Jan;24(1):198-208 [FREE Full text] [doi: [10.1093/jamia/ocw042](https://doi.org/10.1093/jamia/ocw042)] [Medline: [27189013](https://pubmed.ncbi.nlm.nih.gov/27189013/)]
28. Falconer N, Abdel-Hafez A, Scott IA, Marxen S, Canaris S, Barras M. Systematic review of machine learning models for personalised dosing of heparin. *Br J Clin Pharmacol* 2021 Nov 14;87(11):4124-4139. [doi: [10.1111/bcp.14852](https://doi.org/10.1111/bcp.14852)] [Medline: [33835524](https://pubmed.ncbi.nlm.nih.gov/33835524/)]
29. Ghassemi M, Alhanai T, Westover M, Mark R, Nemati S. Personalized medication dosing using volatile data streams. In: Proceedings of the The Workshops of the Thirty-Second AAAI Conference on Artificial Intelligence. 2018 Presented at: The Workshops of the Thirty-Second AAAI Conference on Artificial Intelligence; Feb 2–7, 2018; New Orleans, Louisiana, USA.
30. Ghassemi MM, Richter SE, Eche IM, Chen TW, Danziger J, Celi LA. A data-driven approach to optimized medication dosing: a focus on heparin. *Intensive Care Med* 2014 Sep 5;40(9):1332-1339 [FREE Full text] [doi: [10.1007/s00134-014-3406-5](https://doi.org/10.1007/s00134-014-3406-5)] [Medline: [25091788](https://pubmed.ncbi.nlm.nih.gov/25091788/)]
31. Smith BP, Ward RA, Brier ME. Prediction of anticoagulation during hemodialysis by population kinetics and an artificial neural network. *Artif Organs* 1998 Sep;22(9):731-739. [doi: [10.1046/j.1525-1594.1998.06101.x](https://doi.org/10.1046/j.1525-1594.1998.06101.x)] [Medline: [9754457](https://pubmed.ncbi.nlm.nih.gov/9754457/)]
32. Su L, Liu C, Li D, He J, Zheng F, Jiang H, et al. Toward optimal heparin dosing by comparing multiple machine learning methods: retrospective study. *JMIR Med Inform* 2020 Jun 22;8(6):e17648 [FREE Full text] [doi: [10.2196/17648](https://doi.org/10.2196/17648)] [Medline: [32568089](https://pubmed.ncbi.nlm.nih.gov/32568089/)]
33. Lin R, Stanley MD, Ghassemi MM, Nemati S. A deep deterministic policy gradient approach to medication dosing and surveillance in the ICU. *Annu Int Conf IEEE Eng Med Biol Soc* 2018 Jul;2018:4927-4931 [FREE Full text] [doi: [10.1109/EMBC.2018.8513203](https://doi.org/10.1109/EMBC.2018.8513203)] [Medline: [30441448](https://pubmed.ncbi.nlm.nih.gov/30441448/)]
34. Kong N, Liu X, Liu C, Lian J, Wang H. Deep architecture for Heparin dosage prediction during continuous renal replacement therapy. In: Proceedings of the 2017 36th Chinese Control Conference (CCC). 2017 Presented at: 2017 36th Chinese Control Conference (CCC); Jul 26-28, 2017; Dalian, China. [doi: [10.23919/chicc.2017.8029139](https://doi.org/10.23919/chicc.2017.8029139)]
35. Lian J, Zhao Q. Prediction of heparin dose during continuous renal replacement therapy surgery by using the gradient boosting regression model. In: Proceedings of the 6th International Conference on Control, Decision and Information Technologies (CoDIT). 2019 Presented at: 6th International Conference on Control, Decision and Information Technologies (CoDIT); Apr 23-26, 2019; Paris, France. [doi: [10.1109/codit.2019.8820648](https://doi.org/10.1109/codit.2019.8820648)]
36. Nemati S, Ghassemi MM, Clifford GD. Optimal medication dosing from suboptimal clinical examples: a deep reinforcement learning approach. *Annu Int Conf IEEE Eng Med Biol Soc* 2016 Aug;2016:2978-2981. [doi: [10.1109/EMBC.2016.7591355](https://doi.org/10.1109/EMBC.2016.7591355)] [Medline: [28268938](https://pubmed.ncbi.nlm.nih.gov/28268938/)]
37. Li D, Gao J, Hong N, Wang H, Su L, Liu C, et al. A clinical prediction model to predict heparin treatment outcomes and provide dosage recommendations: development and validation study. *J Med Internet Res* 2021 May 20;23(5):e27118 [FREE Full text] [doi: [10.2196/27118](https://doi.org/10.2196/27118)] [Medline: [34014171](https://pubmed.ncbi.nlm.nih.gov/34014171/)]
38. Bussey H, Francis JL, Heparin Consensus Group. Heparin overview and issues. *Pharmacotherapy* 2004 Aug;24(8 Pt 2):103S-107S. [doi: [10.1592/phco.24.12.103s.36109](https://doi.org/10.1592/phco.24.12.103s.36109)] [Medline: [15334855](https://pubmed.ncbi.nlm.nih.gov/15334855/)]
39. Hirsh J, van Aken WG, Gallus AS, Dollery CT, Cade JF, Yung WL. Heparin kinetics in venous thrombosis and pulmonary embolism. *Circulation* 1976 Apr;53(4):691-695. [doi: [10.1161/01.cir.53.4.691](https://doi.org/10.1161/01.cir.53.4.691)] [Medline: [1253392](https://pubmed.ncbi.nlm.nih.gov/1253392/)]
40. Yeo I. A new family of power transformations to improve normality or symmetry. *Biometrika* 2000 Dec 01;87(4):954-959. [doi: [10.1093/biomet/87.4.954](https://doi.org/10.1093/biomet/87.4.954)]
41. Liu W, Chawla S. A quadratic mean based supervised learning model for managing data skewness. In: Proceedings of the 2011 SIAM International Conference on Data Mining (SDM). 2011 Presented at: 2011 SIAM International Conference on Data Mining (SDM); Apr 28-30, 2011; Mesa, Arizona, USA. [doi: [10.1137/1.9781611972818.17](https://doi.org/10.1137/1.9781611972818.17)]
42. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation. 2016 Presented at: 12th USENIX conference on Operating Systems Design and Implementation; Nov 2 - 4, 2016; Savannah GA USA.
43. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International

- Conference on Knowledge Discovery and Data Mining; Aug 13 - 17, 2016; San Francisco California USA. [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]
44. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4 - 9, 2017; Long Beach California USA.
 45. Nelder JA, Wedderburn RW. Generalized linear models. *J Royal Stat Soc Series A (General)* 1972;135(3):370. [doi: [10.2307/2344614](https://doi.org/10.2307/2344614)]
 46. Friedman JH, Popescu BE. Predictive learning via rule ensembles. *Ann Appl Stat* 2008 Sep 1;2(3):916-954. [doi: [10.1214/07-aos148](https://doi.org/10.1214/07-aos148)]
 47. McMahan H, Holt G, Sculley D, Young M, Ebner D, Grady J, et al. Ad click prediction: a view from the trenches. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013 Presented at: KDD' 13: The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 11 - 14, 2013; Chicago Illinois USA. [doi: [10.1145/2487575.2488200](https://doi.org/10.1145/2487575.2488200)]
 48. Hall P, Kurka M, Bartz A, Haq Z. Using H2O Driverless AI. Mountain View, CA: H2O.ai; 2017.
 49. Rifkin R, Lippert R. Notes on regularized least squares. Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory. 2007 May 1. URL: <https://dspace.mit.edu/handle/1721.1/37318> [accessed 2022-03-21]
 50. Kim S, Koh K, Lustig M, Boyd S, Gorinevsky D. An interior-point method for large-scale -regularized least squares. *IEEE J Sel Top Signal Process* 2007 Dec;1(4):606-617. [doi: [10.1109/jstsp.2007.910971](https://doi.org/10.1109/jstsp.2007.910971)]
 51. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Soft* 2010;33(1):1-22. [doi: [10.18637/jss.v033.i01](https://doi.org/10.18637/jss.v033.i01)]
 52. Wu X, Zhou Z. A unified view of multi-label performance measures. arXiv 2017 Sep 1 [FREE Full text]
 53. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995 Sep;20(3):273-297. [doi: [10.1007/bf00994018](https://doi.org/10.1007/bf00994018)]
 54. Crammer K, Singer J. On the algorithmic implementation of multiclass kernel-based vector machines. *J Mach Learn Res* 2002;2:265-292.
 55. Archer L, Snell KI, Ensor J, Hudda MT, Collins GS, Riley RD. Minimum sample size for external validation of a clinical prediction model with a continuous outcome. *Stat Med* 2021 Jan 15;40(1):133-146. [doi: [10.1002/sim.8766](https://doi.org/10.1002/sim.8766)] [Medline: [33150684](https://pubmed.ncbi.nlm.nih.gov/33150684/)]
 56. Tao H, Li Q, Zhou Q, Chen J, Fu B, Wang J, CLIATHVR multicenter clinical study team. A prediction study of warfarin individual stable dose after mechanical heart valve replacement: adaptive neural-fuzzy inference system prediction. *BMC Surg* 2018 Feb 15;18(1):10 [FREE Full text] [doi: [10.1186/s12893-018-0343-1](https://doi.org/10.1186/s12893-018-0343-1)] [Medline: [29448930](https://pubmed.ncbi.nlm.nih.gov/29448930/)]
 57. Ma Z, Wang P, Gao Z, Wang R, Khalighi K. Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose. *PLoS One* 2018;13(10):e0205872 [FREE Full text] [doi: [10.1371/journal.pone.0205872](https://doi.org/10.1371/journal.pone.0205872)] [Medline: [30339708](https://pubmed.ncbi.nlm.nih.gov/30339708/)]
 58. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ* 2020 Mar 20;368:l6927 [FREE Full text] [doi: [10.1136/bmj.l6927](https://doi.org/10.1136/bmj.l6927)] [Medline: [32198138](https://pubmed.ncbi.nlm.nih.gov/32198138/)]
 59. Scott IA. Machine learning and evidence-based medicine. *Ann Intern Med* 2018 May 01;169(1):44. [doi: [10.7326/m18-0115](https://doi.org/10.7326/m18-0115)]
 60. Sohrahi MK, Tajik A. Multi-objective feature selection for warfarin dose prediction. *Comput Biol Chem* 2017 Aug;69:126-133. [doi: [10.1016/j.combiolchem.2017.06.002](https://doi.org/10.1016/j.combiolchem.2017.06.002)] [Medline: [28689108](https://pubmed.ncbi.nlm.nih.gov/28689108/)]
 61. Pavani A, Naushad SM, Kumar RM, Srinath M, Malempati AR, Kutala VK. Artificial neural network-based pharmacogenomic algorithm for warfarin dose optimization. *Pharmacogenomics* 2016;17(2):121-131. [doi: [10.2217/pgs.15.161](https://doi.org/10.2217/pgs.15.161)] [Medline: [26666467](https://pubmed.ncbi.nlm.nih.gov/26666467/)]
 62. Shifrin M, Siegelmann H. Near-optimal insulin treatment for diabetes patients: a machine learning approach. *Artif Intell Med* 2020 Jul;107:101917. [doi: [10.1016/j.artmed.2020.101917](https://doi.org/10.1016/j.artmed.2020.101917)] [Medline: [32828456](https://pubmed.ncbi.nlm.nih.gov/32828456/)]
 63. Hu Y, Tai C, Tsai C, Huang M. Improvement of adequate digoxin dosage: an application of machine learning approach. *J Healthc Eng* 2018;2018:3948245 [FREE Full text] [doi: [10.1155/2018/3948245](https://doi.org/10.1155/2018/3948245)] [Medline: [30210752](https://pubmed.ncbi.nlm.nih.gov/30210752/)]
 64. Adam G, Rampášek L, Safikhani Z, Smirnov P, Haibe-Kains B, Goldenberg A. Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ Precis Oncol* 2020 Jun 15;4(1):19 [FREE Full text] [doi: [10.1038/s41698-020-0122-1](https://doi.org/10.1038/s41698-020-0122-1)] [Medline: [32566759](https://pubmed.ncbi.nlm.nih.gov/32566759/)]
 65. Scott IA, Abdel-Hafez A, Barras M, Canaris S. What is needed to mainstream artificial intelligence in health care? *Aust Health Rev* 2021;45(5):591. [doi: [10.1071/ah21034](https://doi.org/10.1071/ah21034)]
 66. Scott IA. Hope, hype and harms of Big Data. *Intern Med J* 2019 Jan 24;49(1):126-129. [doi: [10.1111/imj.14172](https://doi.org/10.1111/imj.14172)] [Medline: [30680901](https://pubmed.ncbi.nlm.nih.gov/30680901/)]
 67. Wong A, Otlis E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med* 2021 Aug 01;181(8):1065-1070 [FREE Full text] [doi: [10.1001/jamainternmed.2021.2626](https://doi.org/10.1001/jamainternmed.2021.2626)] [Medline: [34152373](https://pubmed.ncbi.nlm.nih.gov/34152373/)]
 68. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med* 2019 Sep;25(9):1337-1340. [doi: [10.1038/s41591-019-0548-6](https://doi.org/10.1038/s41591-019-0548-6)] [Medline: [31427808](https://pubmed.ncbi.nlm.nih.gov/31427808/)]

69. Sandhu S, Lin AL, Brajer N, Sperling J, Ratliff W, Bedoya AD, et al. Integrating a machine learning system into clinical workflows: qualitative study. *J Med Internet Res* 2020 Nov 19;22(11):e22421 [FREE Full text] [doi: [10.2196/22421](https://doi.org/10.2196/22421)] [Medline: [33211015](https://pubmed.ncbi.nlm.nih.gov/33211015/)]
70. Ghassemi M, Naumann T, Schulam P, Beam A, Chen I, Ranganath R. A review of challenges and opportunities in machine learning for health. *AMIA Jt Summits Transl Sci Proc* 2020;2020:191-200 [FREE Full text] [Medline: [32477638](https://pubmed.ncbi.nlm.nih.gov/32477638/)]

Abbreviations

aPTT: activated partial thromboplastin time
AUC: area under the receiver operating characteristic curve
EHR: electronic health record
ICU: intensive care unit
ML: machine learning
MSH: Metro South Health
RMSE: root mean square error
TR: therapeutic range
UFH: unfractionated heparin

Edited by T Leung; submitted 28.10.21; peer-reviewed by E Sükei, B Hug; comments to author 13.02.22; revised version received 10.04.22; accepted 23.06.22; published 19.09.22

Please cite as:

*Abdel-Hafez A, Scott IA, Falconer N, Canaris S, Bonilla O, Marxen S, Van Garderen A, Barras M
Predicting Therapeutic Response to Unfractionated Heparin Therapy: Machine Learning Approach
Interact J Med Res 2022;11(2):e34533
URL: <https://www.i-jmr.org/2022/2/e34533>
doi: [10.2196/34533](https://doi.org/10.2196/34533)
PMID: [35993617](https://pubmed.ncbi.nlm.nih.gov/35993617/)*

©Ahmad Abdel-Hafez, Ian A Scott, Nazanin Falconer, Stephen Canaris, Oscar Bonilla, Sven Marxen, Aaron Van Garderen, Michael Barras. Originally published in the Interactive Journal of Medical Research (<https://www.i-jmr.org/>), 19.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.i-jmr.org/>, as well as this copyright and license information must be included.